On prediction of regulatory genes by analysis of C.elegans functional networks

O.V. Valba[1,2], S.K. Nechaev[1,3], O. Vasieva[4]

[1]*LPTMS, Université Paris Sud, 91405 Orsay Cedex, France*

[2]*Moscow Institute of Physics and Technology, 141700, Dolgoprudny, Russia*

[3]*P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991, Moscow, Russia*

[4]*Institute of Integrative Biology, University of Liverpool, Crown St., Liverpool, L69 7ZB, UK*

Connectivity networks have recently become widely used in biology due to increasing amounts of information on the physical and functional links between individual proteins. This connectivity data provides valuable material for expanding our knowledge far beyond the experimentally validated via mathematical analysis and theoretical predictions of new functional interactions. In this paper we demonstrate an application of several algorithms developed for the ranking of potential gene-expression regulators within the context of an associated network. We analyze how different types of connectivity between genes and proteins affect the topology of the integral C.elegans functional network and thereby validate algorithmic performance. We demonstrate the possible definition of co-expression gene clusters within a network context from their specific motif distribution signatures. We also show that the method based on the shortest path function (SPF) applied to gene interactions sub-network of the co-expression gene cluster, efficiently predicts novel regulatory transcription factors (TFs). Simultaneous application of other methods, including only interactions with neighborhood genes, allows rapid ranking of potential regulators that could be functionally linked with the group of co-expressed genes. Predicting functions of regulators for a cluster of ribosomal/mRNA metabolic genes we highlight a role of mRNA translation and decay in a longevity of organisms.

I. INTRODUCTION

Analysis of functional properties of biological systems requires the integration of both the functions of their individual components and the properties of the components interactions. Reconstruction of functional networks from the known pair-wise connectivity of biological molecules offers systems level insights into complex biologic phenomena. Though widely used and taken for granted the network-based approach requires very accurate choice of data to reflect a proper balance between maximizing connectivity on one hand, and preserving the reliability of network behavior on the other.

Direct genetic and regulatory interactions as well as indirect indicators of functional links between proteins, such as co-expression, co-occurrence, gene-fusion, *etc*, are used to provide a generic topology of a network. It is known that subnetworks based on genetic and physical protein-protein interactions overlap only weakly [1], which perhaps indicates

different levels of functional organization within complex biological systems. However, it has been also documented that indirect indications or proxies of functional relevance between proteins, such as gene co-expression and genome co-localization are largely complementary and correlate well with ontology-based protein groupings [2, 3]. It is not clear to what extent connectivities of different types affect the general architecture of an integrated network structure.

The gaps in experimentally-derived knowledge on regulatory and structural features of biological systems can be filled to some extent by theoretical predictions. However, a greater understanding of principles of connectivity between biological functions at different levels of cellular organization is needed to avoid false extrapolations in proposed network structures or topologies. There is thus an immediate need for a means of evaluating the generally used organism-specific functional networks and understanding of their underlying properties. Here we show how the statistical analysis of network connectivity can be used to predict new gene expression regulators. We suggest that co-expression clusters can be easily identified as highly-connected islands within the integrated network, and that these islands can be used to suggest new regulatory genes for subsequent verification. For this, we propose a new application of modified statistical algorithm [4], based on so-called "shortest path function" (SPF) to rank the nodes that have most effect on genes expression. This can be also applied to any explicitly defined group of genes.

We retrieved gene co-expression clusters from existing large-scale microarray data on heat shock responses in C.elegans [5] and projected them onto subnetworks composed from a combination of different connectivity types (gene interactions, protein interactions, gene co-expression) in WormNet database of pair-wise undirected functional links between proteins[6]. Correlation between experimentally observed co-expression links and connectivity of genes in Wormbase [7] is shown to be very strong especially for ubiquitously expressed ribosomal, proteosomal and exosomal gene clusters. Via application of SPF method we also predicted several transcription factors (TFs) as potential regulators of the analyzed groups of co-expressed genes. By searching for the most connected regulatory nodes within the neighborhood of each cluster we also noticed a strong link between nonsense mediated decay (NMD) as well as longevity-related genes and the highly connected clusters mentioned above. We suggest that these functional connections may explain a dependence of adult life span on the metabolism of polyamines. Interpretation of the organism-specific integral biologica networks and prediction of protein complexes and genetic regulators from a network context may benefit greatly from our study and the new algorithms.

II. METHODS

A. Data preparation

Microarray data are adopted from [8] where two parental C.elegans strains N2 (Bristol) and CB4856(Hawaii) and the recombinant inbred strains were used to measure gene

expression in a control conditions and after a heat shock. N2 and CB4856, represent two genetic and ecological extremes of C. elegans [9–11]. Their genetic distance amounts to about one polymorphism per 873 base pairs [12]. Both strains have contrasting behavioral phenotypes (solitary versus gregarious) [11] and differ strikingly in their response to a temperature change [5]. In [13] the strains were exposed to 16 ^{0}C and 24 ^{0}C, temperatures that are known to strongly affect phenotypic characteristics such as body size, lifespan, and reproduction [5, 14].

Gene expression patterns were assessed by oligonucleotide microarray hybridization. All microarray data have been deposited by the authors in NCBI's Gene Expression Omnibus (GEO, [15]) and are accessible through the GEO Series accession number listed under the Accession Numbers GSE5395. Absolute expression values have been used for KMC clustering analysis by means of Mev4 application [16] with requested 50 clusters/Euclidian distance. The resultant clusters were used as "co-expression" clusters in our analysis.

String software [17, 18] has been used to reconstruct graphical networks from the sets of C.elegans genes. WormBase database [7] has been used for ID retrieval and translation of IDs into gene names, associated functional annotations and ontological categories. WormNet [6] has been used a source of information on pair-wise interactions between genes. We also used it to retrieve data on separated genetic interactions of C.elegans genes and co-expression links in C.elegans.

B. Methods of statistical analysis of network connectivity

To investigate the connectivity properties of the gene clusters on WormNet and to establish potential regulators for the gene clusters we use the following approach. Denote by $d_{i,M}$ the shortest path along the network from a given vertex (regulatory gene) M to some other vertex (cluster gene) i. Consider the shortest path function (SPF) determined as follows

$$k_M^{SPF} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{d_{i,M}} \tag{1}$$

where the summation is performed over all cluster genes (N is the number of the genes in the cluster, i.e. the cluster size). The connectivity of the gene cluster on the whole network and/or on its subnetworks we describe by SPF defined for all cluster genes

$$k_{cl}^{SPF} = \frac{2}{N(N-1)}\sum_{i,j=1}^{N}\frac{1}{d_{i,j}} \tag{2}$$

here N is the number of the genes in the cluster, $d_{i,j}$ is the shortest path between the nodes i and j. Thus defined, the SPF has a very transparent meaning, since it gives the averaged reciprocal paths between pair of cluster genes. If i and j are not linked on the network, the contribution to SPF from this pair (i, j) equals to zero, while the maximal contribution is reached for directly linked pairs. So, SPF can be used to characterize quantitatively

the connectivity of a gene cluster in a given network. For comparison we use also another method for determination the cluster connectivity. The so-called "connectivity coefficient" is defined as a ratio between the number of *inner* links (which connect only cluster genes) to the number of *all* links which the cluster has in the network:

$$f_{cl} = \frac{N_{in}}{N_{all}} \tag{3}$$

We consider WormNet and its parts as unweighted undirected network. Note that SPF analysis can be easily extended for weighted and directed networks as well.

Another approach to characterize a connectivity (and topology) of a subnetwork (cluster) is to investigate its motifs distribution. A network is a set of vertices (or nodes) and connections between them (links or edges) without self–connections and multiple edges. The network is random if any link occurs independently on others with a certain probability. The network is directed if any link either has an orientation $i \to j$, or is bidirectional $i \leftrightarrow j$. Otherwise the network is non-directed. The network is set by the adjacency matrix. The local topological properties of networks, both directed and non-directed, for given number of vertices and vertex degree distribution are characterized by the rates of connected subgraphs. Since the number of such subgraphs grows combinatorially with their sizes, usually only small subgraphs are considered. In particular, in the works [19–21] only subgraphs of size 3 and 4 have been analyzed for directed and undirected networks.

The rates of subgraphs in a given network depend on the vertex degree distribution. This complicates the comparison of networks of different sizes and of different degree distributions by the rates of their subgraphs. In order to compensate these differences, the procedure of so-called network randomization was proposed in works [20, 21]. In this procedure the network experiences multiple permutations of links under the condition of conservation in each vertex the number of incoming, outgoing and bidirectional links. Using this method an ensemble of randomized versions of a given network is generated, and for every subgraph the statistical significance

$$Z_k = \frac{N_k - \langle N_k \rangle}{\sigma_k}; \quad k = [1...m] \tag{4}$$

is calculated, where N_k is the amount of k-th subgraphs in the initial network; and $\langle N_k \rangle$ and σ_k correspondingly the mean and the standard deviation of subgraphs of given type in the randomized networks; m is the total number of considered subgraphs. Subgraphs with the statistical significance essentially exceeding 1 are called motifs [19–21]. The distribution of motifs in the network under consideration is characterized by the significance profile which is a normalized vector $x = \{x_1, ..., x_m\}$ of statistical significance of all subgraphs of given size. The components of the vector x are

$$x_k = \frac{Z_k}{\sqrt{\sum_{k=1}^{m} Z_k^2}}; \quad k = [1...m] \tag{5}$$

C. Prediction of potential regulators for the gene clusters

We use the SPF coefficients k_M^{SPF}, determined in (1) for a search for potential regulators of co-expressed gene clusters. Note, that potential regulators can belong to the cluster or lie outside of it. For the search of the potential cluster regulators it is important that the regulator is specific to it. It has many connections to the gene cluster but at the same time is weakly connected to other genes in the network, which the cluster belongs to. In other words, the potential regulator must have enough links to control the cluster not being herewith the hub of the network. We introduce the value which takes into account the connectivity of the regulatory gene in the whole network:

$$K_M^{SPF} = \frac{\sum_{i=1}^{N} \frac{1}{d_{i,M}}}{\sum_{i=1}^{G} \frac{1}{d_{i,M}}} \tag{6}$$

By definition, the K_M^{SPF} is the ratio of the sum of the shortest paths from regulator to the cluster genes, and the sum of the shortest path from regulator to all genes in the network. one sees, that K_M^{SPF} measures how specific is the regulator to the cluster. The function (1) can be considered as a first term in expansion of the function (6). We use this value for the search of potential regulators to the gene clusters in the network and its subparts. We compare this method with some other described below.

Namely, for any potential regulator we define the fraction of the cluster genes, which are connected to it:

$$f_M = \frac{N_M}{N} \tag{7}$$

Analogously to (6) we introduce the value, which takes into account the "interaction" of the potential regulator with outer part of the cluster genes:

$$F_M = \frac{N_M}{N_M(G)} \tag{8}$$

Here, $N_M(G)$ is the number of all links, which the potential regulator M has in the whole network G.

III. RESULTS

A. Statistical properties of co-expression clusters

Our first goal is the investigation of the properties of extracted gene clusters in the whole network WormNet. The Fig. 1 presents the dependencies of the cluster's connectivity and SPF on the cluster's size (the number of genes included in the cluster), as well as the motif's

distribution for some clusters. Connectivity coefficient is generally proportional to a cluster size. In average, an expression cluster is composed by genes connected in WormNet stronger than any group of randomly selected genes – see Fig. 1A. The SPF values for all gene clusters lie much higher than the SPF coefficients for random clusters shown in the Fig. 1B (here and below: a random cluster is a cluster with randomly chosen nodes in a WormNet). The most connected clusters are bounded mainly by co-expression links in the WormBase Fig. 1D. We have selected the clusters with very high connectivity coefficients. Such clusters are well-connected and are depicted by red points in the graphs. These clusters also are characterized by different motif's distribution with a prevail of linear 4-nodes chains Fig. 1C. According to the motif's distribution of gene clusters, we can divide them into two groups. The first group includes the gene clusters, in which the fully connected subgraphs dominate and all well-connected clusters belong to this group. The second group is formed by the clusters with a small number of fully connected motifs. It is worth noting, that protein structure networks are also characterized by the same motif distribution [20].

The number of co-expression and physical connections strongly prevail over genetic interactions in the WormBase, which emphasizes the necessity in additional experimental analysis, or in-situ prediction of potential regulatory modules for C.elegans. Our analysis shows that the motif's profiles for expression sub-network of gene clusters are very close to the profiles in the whole WormNet Fig. 1C,F. It means that in our clusters the co-expression links prevail over other links. This observation provides a proof that genes of well-connected clusters are mainly connected to each other by co-expression. Also, our analysis demonstrates that expression links are responsible for the existence of the cycles in well-connected clusters Fig. 1F.

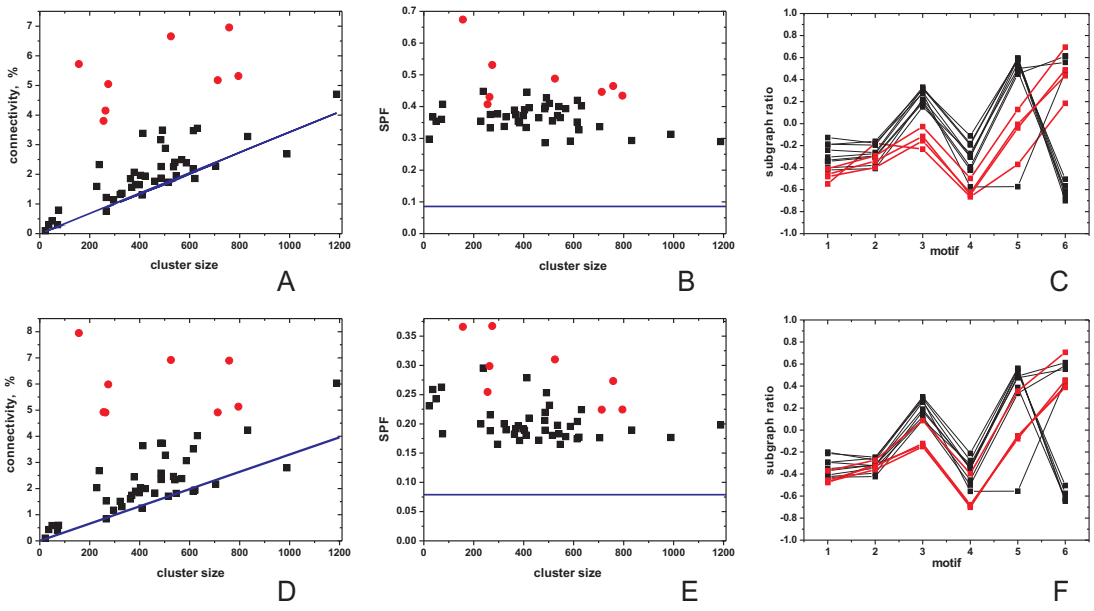

Figure 1: The connectivity coefficient (A and D), SPF (B and E) and motif distribution (C and F) in the whole WormNet (up) and in its expression subpart (below).

B. Prediction of expression cluster regulators

We have used the most connected cluster 1 to find an optimal algorithm to predict potential regulators of mRNA pool for the groups of co-expressed genes. Two different methods have been tested.

The first method is based on ranking the network nodes according to a fraction of cluster genes (FCG) they are directly connected to. This method requires dense connectivity matrix to be efficient. The predicted functions are expected to have a strong and specific involvement in regulation of a particular co-expressed gene cluster. However, the ranked gene list needs filtering to eliminate a malfunctioning.

The second method is based on ranking nodes according to the shortest average distance to all cluster genes (SPF method). As in the first method, the obtained gene lists needs filtering. From the Fig. 1 one can see that the clusters have different topologies in different connectivity subnetworks, with the genetic interactions subnetwork being the least connected. The power of the SPF method is that it compensates the absence of knowledge about regulation of many genes and quantitatively validates the probability of co-regulation of a group of genes by different nodes. SPF method does not require a matrix to be dense and can be applied to a subnetwork of genetic interactions producing the list of regulators without filtering. FCG and SPF methods are applied to a total network and to the co-expression connectivity subnetwork with similar outcomes. The top of the predicted regulators is presented in Table I.

Several predicted regulators are well integrated both in the top cluster networks and in the network reconstructed from the genes with a longevity phenotype retrieved from WormBase. Among the most promising predicted regulators that could connect the top clusters with a longevity regulation are: daf-2, iff-1, cgh-1, tin9.2, car-1. They all are related to mRNA processing/translation/decay and are in a cross-talking relationship (Table I). According to their position in the network, they may play a role of linkers between top connected co-expression clusters related to ribosomal biogenesis, proteosome and central metabolic functions (Fig 1). The SPF method has also been applied to the gene regulatory network. The results of this analysis (the top ranked regulatory nodes) are summarized in the Table II. Note that the SPF method allows us to predict some regulators which we could not detect by the FCG method, and the power of this method had been enforced by its application to a gene regulatory connectivity subnetwork, as it is clearly demonstrated in Table II. We could identify a number of TFs that may be considered for a role of transcription regulators of genes in the top co-expression cluster 1, such as: F30F8.8, transcription initiation factor TFIID subunit 5; R74.3, xbp-1, heat-shock transcription factor; F02E9.4, sin-3-histone deacetylase subunit; 33A8.1, pre-mRNA-splicing factor CWC22. This method also greatly increased the ranking position of daf-2 and genes upstream daf-2 (C25A1.10) or being directly affected by daf-2 mutation (C05C8.3) an immediate potential connection to a group of genes with longevity phenotype that have a strong overlap with our cluster 1.

Here we do not consider exclusion of the hub-regulators by the procedures (6) for SPF method and (8) for FCG method. Our statistical analysis demonstrates that the most

Seq. IDs	**Gene**	**Function**
F57B9.6	inf-1	Transl.initiation/ RNA transport
T05G5.10	iff-1	Transl.initiation/ NMD
Y71G12B.8	Y71G12B.8	RNA helicase/ RNA transport
T10C6.14, T10C6.12, T10C6.11, F45F2.3, F45F2.4, F45F2.12, ZK131.4, ZK131.6, ZK131.8, ZK131.10, K06C4.10, K06C4.11, K06C4.4, K06C4.3, K06C4.12, ZK131.1, K06C4.2, F35H10.1, F17E9.12, F17E9.13, C50F4.7, K03A1.6, C50F4.5, F08G2.2, B0035.9, B0035.7, F07B7.9, F07B7.10, F07B7.4, F07B7.3, F07B7.11, F54E12.3, F54E12.5, F55G1.11, F55G1.10, F22B3.1, H02I12.7, T23D8.5, T23D8.6	38 His genes	Histones
C41D11.2	eif-3.H	Transl.initiation
F32E10.1	nol-10	nucleolar protein, polyglut. binding
F54H12.6	eef-1B.1	Elongation factor
C01F6.5	aly-1	RNA export
M163.3	his-24	Histones
B0564.1	tin-9.2	decay/ NMD
Y18D10A.17	car-1	decay/decapping
F56D12.5	vig-1	RISC component/miRNA binding
F26D10.3	hsp-1	splicing
R04A9.4	ife-2	Transl.initiation

Table I: Top of the predictable regulators for test cluster N1 by FCG method

connected components of WormNet are included in the test cluster $No1$, so the methods (6) and (8) give about the same list of potential regulators as FCG and SPF method respectively; the ranging order is slightly changed, for example, in a green frame in the figure we show gene (exc-7) that has been excluded in our SPF-hub-exclusion algorithm because its connectivity exceeds a fixed threshold.

The Fig. 2 illustrates a typical position of the predicted potential regulators for the Cluster 1. Nodes predicted by FCG method (purple frame are proximal to the cluster or even inside the cluster. The nodes predicted by SPF method may be significantly distant from the many nodes in the cluster (ces-1, eor-1, orange frames on Fig. 2). Though the connections between the SPF-predicted node and the cluster may include several intermediate steps, the majority of these steps do contain the nodes that can translate signals at the level of mRNA pool regulation, potentially representing complexes of proteins with a joint regulatory performance. As one sees, the type of connectors utilized by different algorithms differs essentially: experimental, regulatory connections are fundamental for SPF and more dense, co-expression ones, for FCG.

Seq. IDs	**Gene**	**Function**
Y55D5A.5,B0334.8,Y116F11B.1	daf-2, age1, daf-28	insulin/aging
F35H8.5	exc-7	mRNA processing
W10D5.1	mef-2	TF
C17D12.2	unc-75	Splicing
C47G2.2	unc-130	TF
F30F8.8	TFIID	Transl.initiation
R74.3	xbp-1	TF, histone modulation
F33A8.1	cwc22	Splicing
C41C4.4	xre-1	(RNA processing) decay/ processing
C37H5.8	hsp-6	Decay
C26D10.2	hel-1 (helicase)	DNA helicase
C07H6.5	cgh-1 (decapping)	decay/ decapping
F02E9.4	sin-3 (HDAC)	histone modulation
M163.3	his-1	Histone
C25A1.10	dao-5	rRNA transcription/aging
ZC247.3	lin-11	TF
R107.8	lin-12	TF
C05D9.5	ife-4	Transl.initiation
R11E3.6	eor-1	TF
F43G9.11	ces-1	TF
ZK909.4	ces-2	TF

Table II: Top of the predictable regulators for test cluster 1 by SPF method on genetic subnetwork

IV. DISCUSSION

Characterizing the degree of connectivity for a given gene to a specific set of genes in the network as a normalized sum of inverse distances in a network, is very natural and brings us back to the applications of "harmonic means" for graph analysis . The main conjecture behind the application of harmonic mean to networks is as follows; if a vertex has multiple links to other vertices, the information is sent "in parallel", i.e. concurrently along the network. Thus, one can define the "efficiency", $e_{i,j}$ in communication between vertices i and j as the inverse of the shortest distance, i.e. $e_{i,j} = 1/d_{i,j}$, see [22]. The average efficiency is straightforwardly related to the definition of SPF. It should be noted that the interpretation of the SPF function as the efficiency of communication could be very useful in further dynamic analysis of the networks. Actually, let v be the velocity, with which the information travels along the network, then the amount of information sent from the node i to the node j per unit time is just $v/d_{i,j}$. The performance, P, is the total amount of information propagating over the network per unit time [22]. In our forthcoming works we plan to analyze the clusters taking into account their limited speed of information propagation. The concept of performance seems very appropriate for that.

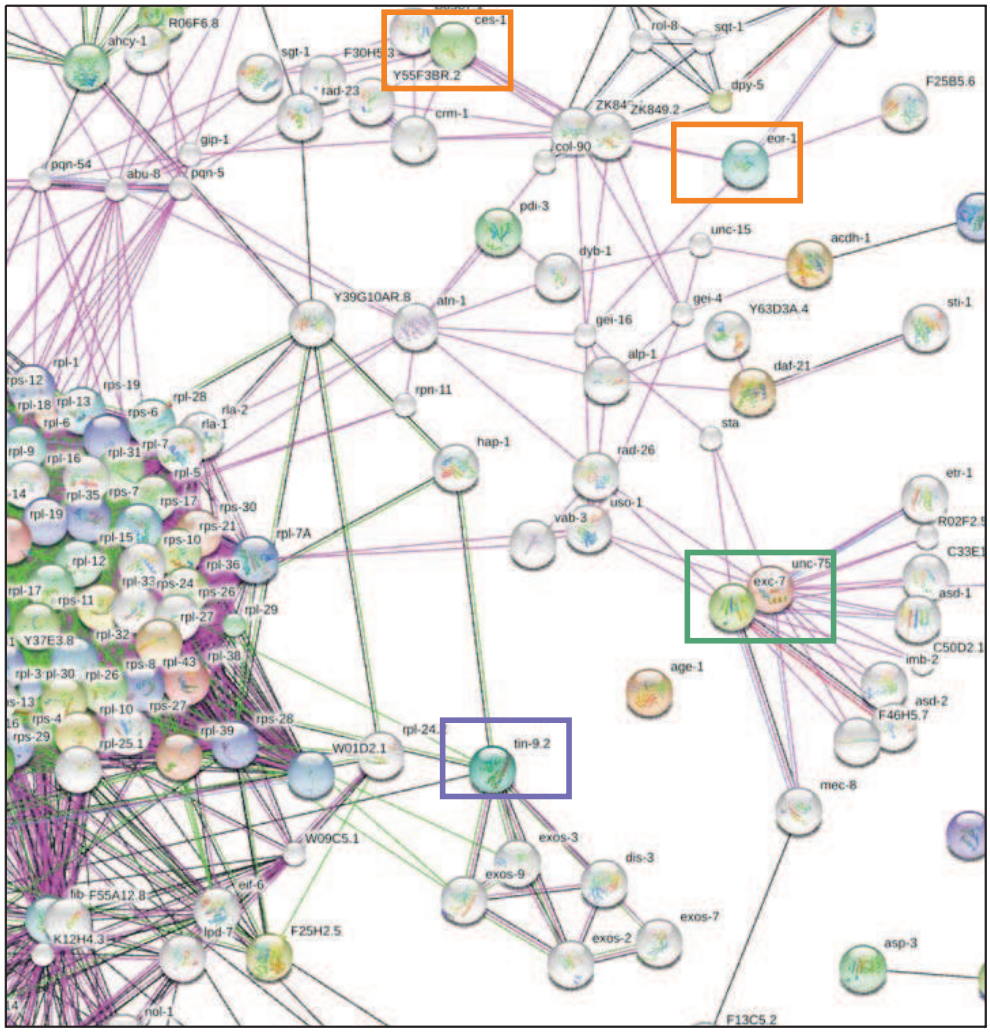

Figure 2: Connectivity between the predicted regulators and Cluster 1 in STRING Network browser. Evidence view for high confidence (0.700) connections. Pink connectors - experimentally derived interactions, black-co-expression, and green-co-localization in genomes, blue-co-occurrences in genomes. Colored circles-input genes, white circles-the most associated additional nodes (set number of 200) automatically added by a STRING software on a request to increase a connectivity between uploaded functions. Predicted potential regulators are shown in frames: orange-SPF method, purple-FCG method, green node excluded in hub-exclusion SPF method.

Well-connected clusters of co-expressed genes described in this paper largely represent protein functional complexes, and they can be distinguished by the presence of a specific well-connected-6 link (unoriented) motif. This highly-connected motif can be used for detection of protein functional complexes (islands) in integral networks. These islands, in turn, serve in prediction of new regulatory nodes. In this study we used gene clusters derived from gene absolute expression values data that probably increase detection of true protein complexes expressed from indeed highly co-regulated genes [23].

Among the most interconnected clusters are the ones for ribosomal proteins and the

regulation of translation, proteosome, respiratory complex 1 and several central metabolic functions. Using the most interconnected cluster 1 we tried to detect potential regulators from the associated network context. Due to a non-directional nature of the edges in WormBase and an absence of directions in the co-expression network we were unable to distinguish between a cause, consequence or undirected physical interaction in a connected pair of proteins/functions and may only suggest the presence of the functional linkage between the expressed genes and the regulators. However, additional data from literature mining will likely help to vectorize the predicted interactions.

The cluster, which we used for the method validation, contains a large number of genes involved in the translational machinery as well as several genes with a central metabolic role, among which we found a large number of genes associated with a longevity phenotype in WarmBase database (Table I). The protein translation processes indeed have been recently considered for a central role in the regulation of aging processes [24, 25] and we assumed that the regulators predicted in this study may also be linked to the processes underlying aging and the control of longevity.

The role of regulation of translational machinery by the insulin pathway in aging has been widely discussed in literature [26, 27], however, the regulatory modules affecting the expression of the related genes downstream of daf-2 have not been clearly defined. The candidates suggested by the FCG algorithm, iff-1 and bir-2 were shown to depend on daf-16-insulin response [28] and iff-1 has also been detected in gene expression screen for the longevity phenotype in C/elegans [25]. iff-1 is a eIF-5A homolog [29], and eIF-5A that links processes of mRNA translation to the nonsense-mediated mRNA decay (NMD) [30]. Activation of eIF-5A requires posttranslational modification of one of the protein's lysines into hypusine, and the enzyme that catalyses the first step of this modification, deoxyhypusine synthase, performs the NAD-dependent oxidative cleavage of spermidine [31]. Spermidine is known to be involved in life span regulation and reproduction in a range of different organisms [32–34], though its mechanism of action is not clear. We suggest that its stimulatory role in NMD via regulation of eIF-5A may be of importance in regulating translation and as a consequence the life span of an organism. Interestingly, the ribosome maturation as well as the mRNA binding SBDS protein [35, 36] that is linked in a network to iff-1 and tin9.2, are both required for the longevity phenotype of daf-2 [37].

Our analysis points to a potential role of mRNA decay processes downstream of the insulin-dependent pathway in regulating translation and longevity. We suggest that elements of translational machinery, that are regulated via insulin/caloric restriction, may be indeed non-responsive to temperature changes as we see on the example of the Cluster 1 genes. Such persistence of expression may require specific mechanisms of adjustment to altered kinetics of biochemical reactions and may indeed involve a regulated mRNA decay process. Homeostasis of pathways regulated by nutrients supply regardless the temperature may lead to a very species-specific dynamic of cell survival and growth and an organism‘s life span adapted to the specific ecological dynamics of nutrients flow. The transcription factors predicted in our study by SPF method may also occur to be involved in aging and regulation of longevity . The genes cgh-1 [38], dao-5 [39], hel-1 [40] were already linked to aging processes downstream

daf-2, daf-16, and in case of dao-5 – to a daf-16 independent pathway associated with determination of the adult life span GO-term in WormNet database. Analysis of genotype-phenotype relationships [41] when more data for the listed genes are available would allow deeper understanding of the direction of the defined links and more narrow prediction of their function.

This work provides some new insights to the structure of biological functional networks and highlights the aspects that need to be considered in prediction of regulatory nodes, protein complexes and regulatory modules from a multilevel network context. We hope that it could be useful in application to analysis of other organism‘s networks and for improvement of analytic methods and software in the relevant fields.

Acknowledgments

The authors are very grateful to Andrew Cossins for useful comments and to L. Basten Snoek and Jan Kammenga for providing us with the experimental data. The work is supported by ERASYSBIO Grant and is a part of the GRAPPLE project.

[1] M. Costanzo et al. The Genetic Landscape of a Cell. Science. 2010; 327:425–431. doi:10.1126/science.1180823.

[2] Mansour TA. S., Markku S., Tommi O. L. and Mauno V. Dynamic covariation between gene expression and proteome characteristics. BMC Bioinformatics. 2005; 6:215. doi:10.1186/1471-2105-6-215.

[3] Fagan A., Culhane AC., Higgins DG. A multivariate analysis approach to the integration of proteomic and gene expression data. Proteomics. 2007; 7(13):2162-71. http://www.ncbi.nlm.nih.gov/pubmed/17549791.

[4] Vohradsky J. Neural model of the genetic network. J. Biol. Chem. 2001; 276 (39): 3616873. doi:10.1074/jbc.M104391200.

[5] Gutteling EW., Riksen JA., Bakker J., Kammenga JE. Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in Caenorhabditis elegans. Heredity. 2007; 98(1):28-37. http://www.ncbi.nlm.nih.gov/pubmed/16955112.

[6] Probabilistic Functional Gene Network of Caenorhabditis elegans: http://www.functionalnet.org/wormnet/.

[7] WormBase: http://www.wormbase.org.

[8] Li Y,Alvarez OA., Gutteling EW., Tijsterman M., Fu J., et al. Mapping Determinants of Gene Expression Plasticity by Genetical Genomics in C. elegans. PLoS Genet. 2006; 2(12):e222. doi:10./journal.pgen.0020222.

[9] Hodgkin J., Doniach T. Natural variation and copulatory plug formation in Caenorhabditis elegans. Genetics. 1997; 146(1):149–64. http://www.ncbi.nlm.nih.gov/pubmed/9136008.

[10] O. Alda Alvarez, T. Jager, S.A.L.M. Kooijman and J.E. Kammenga. Responses to stress of Caenorhabditis elegans populations with different reproductive strategies. Functional Ecology. 2005; 19:656–664. doi:10.1111/j.1365-2435.2005.01012.x.

[11] de Bono M., Bargmann CI. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in C. elegans. Cell. 1998; 94: 679–689. http://www.ncbi.nlm.nih.gov/pubmed/9741632

[12] Wicks SR., Yeh RT., Gish WR., Waterston RH., Plasterk RH. Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. Nat Genet. 2001; 28:160–164. http://www.ncbi.nlm.nih.gov/pubmed/11381264.

[13] Anderson, J. L., L. Albergotti, S. Proulx, C. Peden, R. B. Huey et al. Thermal preference of Caenorhabditis elegans : a null model and empirical tests. J. Exp. Biol. 2007; 210:3107–3116.

[14] Kammenga, J. E., P. C. Phillips, M. de Bono and A. Doroszuk. 2008 Beyond induced mutants: using worms to study natural variation in genetic pathways. Trends Genet. 2008; 24:178–185. doi:10.1016/j.tig.2008.01.001.

[15] Gene Expression Omnibus: http://www.ncbi.nlm.nih.gov/geo.

[16] Saeed AI., Sharov V., White J., Li J., Liang W., Bhagabati N., et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003; 34(2):374-8. http://www.ncbi.nlm.nih.gov/pubmed/12613259.

[17] Szklarczyk D., Franceschini A., Kuhn M., Simonovic M., Roth A., Minguez P., Doerks T., Stark M., Muller J., Bork P., Jensen LJ., von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011; 39(Database issue):D561-8.

[18] STRING - Known and Predicted Protein-Protein Interactions: http://string-db.org/

[19] Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007; 8(6):450-61. http://www.ncbi.nlm.nih.gov/pubmed/17510665.

[20] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U.Alon. Superfamilies of Evolved and Designed Networks. Science. 2004; 303:1538–1542. doi:10.1126/science.1089167.

[21] R. Milo, S. Itzkovitz, N. Kashtan, D. Chklovskii, U.Alon. Network Motifs: Simple Building Blocks of Complex Networks. Science. 2002; 298:824–827. doi:10.1126/science.298.5594.824.

[22] V. Latora and M. Marchiori. Efcient Behavior of Small-World Networks. Phys. Rev. Lett. 2001; 87(19). doi: 10.1103/PhysRevLett.87.198701.

[23] AC. Gavin at al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006; doi:10.1038/nature04532.

[24] Syntichaki and Tavernarakis. Signaling pathways regulating protein synthesis during ageing. Exp Gerontol. 2006; 41(10):1020-5. http://www.ncbi.nlm.nih.gov/pubmed/16829008.

[25] Hamilton et al. A systematic RNAi screen for longevity genes in C. elegans. Genes Dev. 2005; 19(13): 15441555. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1172061.

[26] A. Dillin et al. Timing Requirements for Insulin/IGF-1 Signaling in C. elegans. Science. 2002; doi: 10.1126/science.1074240.

[27] Kenyon C. The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. Philos Trans R Soc Lond B Biol Sci. 2011; 366(1561):9-16. doi: 10.1098/rstb.2010.0276.

[28] Murphy CT. et al. Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. Nature. 2003; 424(6946):277-83. http://www.ncbi.nlm.nih.gov/pubmed/12845331.

[29] Hanazawa M, Kawasaki I, Kunitomo H, Gengyo-Ando K, Bennett KL, Mitani S, Iino Y. The Caenorhabditis elegans eukaryotic initiation factor 5A homologue, IFF-1, is required for germ cell proliferation, gametogenesis and localization of the P-granule component PGL-1. Mech Dev. 2004;121(3):213–24. http://www.ncbi.nlm.nih.gov/pubmed/15003625.

[30] Schrader R., Young C., Kozian D., Hoffmann R., Lottspeich F.J. Temperature-sensitive eIF5A mutant accumulates transcripts targeted to the nonsense-mediated decay pathway. Biol Chem. 2006; 281(46):35336-46. http://www.ncbi.nlm.nih.gov/pubmed/16987817.

[31] Park MH., Joe YA., Kang KR. Deoxyhypusine synthase activity is essential for cell viability in the yeast Saccharomyces cerevisiae. J Biol Chem. 1998; 273(3):1677–83. doi: 10.1074/jbc.273.3.1677.

[32] M. Kaeberlein. Spermidine surprise for a long life. Nature cell biology. 2009; 11:1277–1278. doi:10.1038/ncb1109-1277.

[33] R. Kaur-Sawhney, LM. Shih,H.E. Flores and A.W. Galston. Relation of Polyamine Synthesis and Titer to Aging and Senescence in Oat Leaves. Plant Physiol. 1982; 69(2): 405410.

[34] Eisenberg T. et al. Induction of autophagy by spermidine promotes longevity. Nat Cell Biol. 2009;11(11):1305-14. doi: 10.1038/ncb1975.

[35] T. F. Menne et al. The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. Nature Genetics. 2007; 39:486–495. doi:10.1038/ng1994.

[36] Vasieva O. Role of Shwachman-Bodian-Diamond syndrome protein in translation machinery and cell chemotaxis: a comparative genomics approach. Adv Appl Bioinform Chem. 2011;4:43-50. doi: 10.2147/AABC.S23510.

[37] Samuelson AV., Carr CE., Ruvkun G. Gene activities that mediate increased life span of C. elegans insulin-like signaling mutants. Genes Dev. 2007;21(22):2976-94. http://www.ncbi.nlm.nih.gov/pubmed/18006689.

[38] Navarro RE, Shim EY, Kohara Y, Singson A, Blackwell TK. cgh-1, a conserved predicted RNA helicase required for gametogenesis and protection from physiological germline apoptosis in C. elegans. 2001;128(17):3221-32. http://www.ncbi.nlm.nih.gov/pubmed/11546739.

[39] Simmer et al. Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol. 2003;1(1):E12.

[40] J.Halascheck-Wiener, et al. Analysis of long-lived C. elegans daf-2 mutants using serial analysis of gene expression. Genome Res. 2005; 15(5): 603–615. doi: 10.1101/gr.3274805.

[41] Lee I., Lehner B., Crombie C., Wong W., Fraser AG., Marcotte EM. A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans. Nat Genet. 2008;40(2):181–8. doi: 10.1038/ng.2007.70.